Citation for published version (APA):
Multi-OMICS Analyses of Frailty and Common Widespread Pain Suggest Involvement of Shared Neurological Pathways

Gregory Livshits\textsuperscript{1,2}, Ida Malkin\textsuperscript{1}, Ruth C.E. Bowyer\textsuperscript{2}, Serena Verdi\textsuperscript{2}, Jordana Bell\textsuperscript{2}, Cristina Menni\textsuperscript{2}, Frances M. K. Williams\textsuperscript{2,\#}, Claire J. Steves\textsuperscript{3,\#}

\textsuperscript{1}Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Israel;

\textsuperscript{2}Department of Twin Research and Genetic Epidemiology, King’s College London, UK;

\textsuperscript{3}Clinical Age Research Unit, King’s College Hospitals Foundation Trust, London, UK.

\#Corresponding authors:
Dr. CJ Steves and Dr. FMK Williams
Department of Twin Research and Genetic Epidemiology
King’s College London, St Thomas’ Hospital
London SE1 7EH, UK
Tel: +44 (0)20 7188. 6743
Fax: +44 (0)20 7188 6761

E-Mail: claire.j.steves@kcl.ac.uk & frances.williams@kcl.ac.uk
Abstract
Common widespread pain (CWP) and frailty are prevalent conditions in older people. We have shown previously that interindividual variation in frailty and CWP is genetically determined. We also reported an association of frailty and CWP caused by shared genetic and common environmental factors. The aim of the present study was to use omic approaches to identify molecular genetic factors underlying the heritability of frailty and its genetic correlation with CWP.

Frailty was quantified through the Rockwood Frailty Index (FI) as a proportion of deficits from 33 binary health deficit questions in 3626 female twins. CWP was assessed using a screening questionnaire. Omics analysis included 305 metabolites and whole-genome (>2.5x10^6 SNPs) and epigenome (~1x10^6 MeDIP-seq regions) assessments performed on fasting blood samples. Using family-based statistical analyses, including path analysis, we examined how FI scores were related to molecular genetic factors and to CWP, taking into account known risk factors such as fat mass and smoking.

FI was significantly correlated with 51 metabolites after correction for multiple testing, with 20 metabolites having P-values between 2.1x10^{-6} and 4.0x10^{-16}. Three metabolites (uridine, C-glycosyl tryptophan, N-acetyl glycine) were statistically independent and thought to exert a direct effect on FI. Epiandrosterone sulphate, previously shown highly inversely associated with CWP, was found to exert an indirect influence on FI. Bioinformatics analysis of GWAS and EWAS showed FI and its covariation with CWP was through genomic regions involved in neurological pathways.

Conclusion
Neurological pathway involvement accounts for the associated conditions of aging CWP and FI.

Key words: frailty, chronic pain, metabolite, GWAS, EWAS, path analysis
Introduction

Age-related loss of physiological function negatively affects quality of life. This deterioration, defined in general as frailty, leads to an increased risk of illness, dependency, and adverse outcomes, including falls, delirium and disability, and death (5). UN reports consistently suggest that populations around the world are aging rapidly (http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf) and that, between 2015 and 2030, the number of people in the world aged ≥60 years is projected to grow by 56% (from 901 million to 1.4 billion). This situation represents one of the most significant and unprecedented social transformations of the twenty-first century, with implications for virtually all sectors of society and in particular health care systems.

The biological mechanisms underlying frailty have been extensively studied in recent years (15,26). We have shown that interindividual variation of frailty scores significantly depends on genetic factors, which explain 34% of their variation. We also found that chronic widespread musculoskeletal pain (CWP) - another prevalent age-related condition - is significantly associated with the Rockwood Frailty Index (FI), and that this association is caused by shared genetic and common environmental factors. Bivariate variance component analysis revealed a strong and significant genetic correlation ($R_G=0.69±0.02$) between the two conditions, after adjustment for age, smoking, and body composition (fat mass).

This raises the important question of the nature of the common genetic factors underlying the two phenotypes. One approach to the problem is the genome-wide association study (GWAS). Recent history of GWAS, however, shows that although making a significant contribution to understanding the biological architecture of many common complex traits, many studies of small sample size demonstrated results having a very low level of replication, containing extensive false-positive findings, and able to explain only a small proportion of the genetic variance (heritability) (2,24,34). Modern molecular technologies
allow simultaneous measurement of thousands of measures of biological molecules from DNA polymorphisms (genomics), DNA methylation often causing genes activation/deactivation (epigenomics), an array of endometabolites (metabolomics), and others. This, in combination with the development of bioinformatics tools, creates the basis for new multi-omics data analysis strategies of a complex phenotype (23,28).

We have reported previously that CWP is strongly associated with steroid hormone metabolism, in particular with epiandrosterone sulphate (EAS) (12). Epigenomewide association study (EWAS) of this sample provided evidence of neurological pathway involvement in CWP (14). Thus, the main aim of the present study was twofold. First, implementing complex genomic, epigenomic, and metabolomic data analysis, we attempted to identify major molecular pathways affecting FI score variation. Second, using comparative statistical and bioinformatics study of the two multiomics outcomes, we attempted to identify common molecular genetic factors involved in the significant genetic correlation between these two conditions.

Materials and Methods

Study design.

Figure 1 illustrate the major steps undertaken in this study. The further description of the material and methods follows this flow chart.

Sample and Study Phenotypes.

We compared simultaneously metabolomics and genomic factors affecting CWP [assessed by us previously (12,14)] and FI (examined in this paper). We described the study sample and the methods of FI assessment in details elsewhere (15). Briefly, participants of this study were 3626 UK female volunteers (with age ranging from 17 to 93 years, with mean 60.5±13.9 years) from the NIHR BRC TwinsUK BioResource. The sample included 1696
monozygotic (MZ) and 1152 dizygotic (DZ) twins, and 778 singletons ascertained from the
general population. Subjects in TwinsUK are sent regular questionnaires for completing and
are invited intermittently to attend for a clinical visit. Each participant was assessed for study
primary phenotypes and potential covariates, including age, smoking, basic anthropometric
measurements, and body composition as assessed by DXA technology. All participants gave
written informed consent. The St. Thomas’ Hospital research ethics committee approved the
project.
Frailty was quantified through the Rockwood FI and was created as a proportion of deficits
from 33 binary health deficit domains (15,22). Originally, assessment of frailty contained a
pain component. In this study, the pain component was omitted, and individual FI- scores
were recalculated correspondingly, to avoid possible duplication with the CWP assessment.
CWP was assessed using the London Fibromyalgia Epidemiology Study Screening
Questionnaire that had been sent to twins for self-completion, without reference to the co-
twin (31), and has been described by us previously (12).

Genomics.
Genotyping was performed in 3 batches on the Illumina Human Hap300 and Human Hap610-
Quad arrays, the results were collated and quality control was performed. Only SNPs with
genotyping rate ≥95% and non-significantly deviating H-W equilibrium, with p≥0.0001, were
retained in the analysis. The 2.5 mln SNP genotype data were available 2286 individuals.
Details of genotyping and quality control were repeatedly previously reported elsewhere (e.g.
32).

Epigenomics.
DNA methylation was profiled across the genome using MeDIP-sequencing followed by
dNA methylation quantifications to assess epigenome variation, as previously described in
these data (13). Briefly, the MeDIP-sequencing protocol resulted in an average of
15,684,723 high quality uniquely mapping reads (BWA) that were subsequently extended to 350 bp to represent the average MeDIP fragment size. Fragments per kilobase per million were quantified in bins (methylation sites) of 500 bp (250 bp overlap) genome wide using MEDIPS v1.6 (4). Methylation levels were finally assessed at 11,524,145 genomic regions of size 500bp (bins), genome wide in each individual in the sample (N=1820). In the epigenome wide association analysis (EWAS), we only considered bins that displayed significant correlation between longitudinal samples available for each individual, where methylation levels were measured at least 3 years apart; and we denoted these bins as lsBINs (13). After quality control, 723,029 lsBINs remained and were considered in the downstream analyses in the study.

Metabolomics.

Metabolon Inc. performed, non-targeted ultrahigh-performance liquid chromatography and mass spectrometry on fasting plasma samples of TwinsUK participants (27). Raw data were median normalized for daily fluctuations of the method and then inverse normalized. In our sample 2530 individuals had metabolic traits. Finally, 305 metabolites having complete data in more than 2000 participants were used in the present study. However, of them 103 metabolites represented unknown biochemical compounds and were excluded from the following analyses, despite highly significant correlations in some instances. Further details are given in our previous study (12).

Statistical analysis and bioinformatics.

Statistical analysis was conducted in the following major steps (Fig 1). First, a series of linear regression analyses was carried out to test for the association between the each of the available metabolites and FI-scores, with simultaneous adjustment for age. Metabolites found to be significantly correlated with FI-scores, after correction for multiple testing by false discovery rate [(FDR=5.0E-04 (1)], were selected for further analysis. To establish the
independent effect of each metabolite, they were simultaneously examined in multiple regression analysis with FI-scores as dependent variable. In addition, we examined the effect of age and relative fat mass (as covariates), taking into account familial structure of the sample using MAN statistical package (http://www.tau.ac.il/~idak/hid_MAN.htm). Next, implementing GenABLE package (http://www.genabel.org/packages/GenABEL), GWAS of the FI and metabolites significantly associated with it were conducted using the entire sample. Since the analysis discovered a number of significant associations between FI and CWP, we implemented path analysis (http://people.exeter.ac.uk/SEGLea/multvar2/pathanal.html) to determine direction of effects. Further explanation is given below.

An epigenome-wide association study (EWAS) was conducted. The sample was divided into two groups: Gr1, including 50 pairs of frailty discordant MZ twins, and Gr2, including all the remaining individuals (N=1720). Twin discordance was determined by intra-pair difference in FI-scores, weighted to pair mean FI score, i.e. $D_j = (FI_{j1} - FI_{j2})/0.5(FI_{j1} + FI_{j2})$. The fifty most discordant twin pairs were selected, and methylation levels compared by a paired t-test. Per bin methylation levels were correlated with FI-scores in the Gr2 sample, or compared by paired t-test between the affected and unaffected on CWP discordant twins and in the remaining sample (N=1720). $P$-values for the results obtained in two non-overlapping samples were combined using Fisher’s method (8).

GWAS and EWAS results were subjected to bioinformatic and annotation analyses, for example, implementing gene ontology (GO) methods (Fig 1). In the epigenome analysis we first assigned IsBINs to the nearest ENSEMBL gene using MEDIPS package for R (11). For genes with multiple bins assigned we retained the IsBIN with the lowest $P$-value for association with CWP. Using Fisher’s approach, we took the combined $P$-values and carried out GO analysis using the weight01 algorithm implemented in the topGO package for
The statistical significance of over-representation of GO terms was estimated using Fisher’s exact test. To make the study comparable with our CWP results (12), two GO domains have been analyzed, Biological Process (BP) and Cellular Component (CC). QIAGEN’s Ingenuity Pathway Analysis (www.qiagen.com/ingenuity) was used for pathway analysis. A similar approach was implemented to summarize and interpret the GWAS data obtained for both FI and CWP.

Results

1. Descriptive statistics.

Present study characteristics of the phenotypes are given in Table 1 with a detailed description of FI given recently elsewhere (15). The preliminary analysis showed that FI-scores were significantly correlated with age ($r=0.446$), BMI ($r=0.326$), relative fat mass, $\text{FAT/H}^2$ ($r=0.330$), all having $p<0.0001$. FI also strongly depended on smoking habits increasing almost linearly between non-smokers, previous smokers and current smokers: 0.216 (SD=0.133), 0.242 (SD=0.138), 0.254 (SD=0.149), $F_{(2d.f.)}=21.8$, $p=4.1*10^{-10}$, adjusted for age. In the following analysis the effect of these covariates were taken into account. Only association with $\text{FAT/H}^2$ remained significant if tested simultaneously with BMI.

2. Metabolomics Analysis with FI

There were 305 chemical compounds for analysis after excluding metabolites that had >20% missing values (10) and with unknown identity. Table 2 shows the 20 most significantly correlated metabolites with FI (p=2.1$^{-06}$ to p=4.0$^{-16}$), which were all significant at $p<0.001$, after FDR correction for multiple testing (1). Variations of these metabolites were not independent and many showed significant inter-correlation with one another. We tested them simultaneously in multivariable regression analysis (Table 3), finding five metabolites showing statistically independent effect (p-values ranged between 3.6$^{-03}$ and 2.4$^{-07}$). These
were EAS, uridine, C-glycosyl tryptophan (C-GT), N-acetyl glycine (N-AG), and indolepropionate. Except C-GT, all metabolite levels decreased with increasing FI-scores, especially EAS (p=3.3\textsuperscript{-06}). C-GT, indole propionate and N-AG are related to protein metabolism, and the first two molecules involved in tryptophan metabolic pathway. However, when the data were examined taking into account familial structure and potential effect of heritability, the results changed slightly (Table 3, two right hand columns), although the general pattern remained the same. The main difference was that indolepropionate was no longer statistically significant.

We conducted metabolic pathway analysis focusing on pathway enrichment analysis (http://www.metaboanalyst.ca/faces/home.xhtml). This analysis was conducted twice: first, the top 20 metabolites (Table 2) were tested. Next, 51 metabolites with p-value \leq 0.0002 (corrected for multiple testing) were examined. In the first analysis, galactose, pyrimidine, arginine and proline and D-glutamine and D-glutamate metabolism pathways were suggested, with p-values 0.014 - 0.046. In the second analysis, metabolic pathways of gluconeogenesis, galactose, taurine and hypotaurine, alanine, and some others were suggested, with p-values 0.002 - 0.033. After FDR correction however no significant metabolic pathways were identified.

3. Genome Wide Association Study (GWAS) of metabolites associated with FI

A multiple linear regression model with additive genetic effect was applied to test for FI-score – genotype association using ~2.5 million genotyped and/or imputed autosomal SNPs. Other covariates adjusted in the model-included age and relative fat mass. In addition, we similarly tested each of the metabolites associated with frailty phenotype. The results are presented as series of Manhattan plots (Figure S1 and Table S1, supplementary material 1, available at http://links.lww.com/PAIN/A639). FI showed no genome-wide significant associations, with top p-values ranging between the 10\textsuperscript{-3} to 10\textsuperscript{-4}. GWAS of the four
significantly associated metabolites showed a different pattern. Three of them, specifically uridine, N-AG and EAS displayed strong association with the single genomic region, with the top p-values correspondingly: \( <10^{-12} \) (Chr#22, mapped to 49304328 - 49318618bp, rs131794), \( <10^{-74} \) (Chr#2 mapped to 27596107 - 27584444bp, rs1260326) and \( <10^{-76} \) (Chr#7 mapped to 98994442 - 99024762bp, rs1581492). For variation of C-GT we found no genome-wide significant associations.

4. Epigenome Wide Association Study (EWAS) of FI

Testing 723,029 lsBINs in 50 FI discordant MZ twin pairs (Gr1) implementing paired t-tests revealed overall \( N_D=27485 \) bins that showed nominally significant associations (p<0.05), and of these, the top 20 association signals were ranged \( p=7.01 \times 10^{-5} - 2.17 \times 10^{-6} \). Correlation analysis of lsBINs with FI-scores in our main sample (Gr2) identified \( N_M=31430 \) nominally significant correlations, with top 20 association \( p = 3.76 \times 10^{-5} - 4.02 \times 10^{-6} \). The results of both analyses were combined by Fisher’s test, which detected 27781 nominally significant results showing the same direction of association in both study subsamples. The 20 top combined results are shown in Table S2, (supplementary material 1, available at http://links.lww.com/PAIN/A639). These data, as well as the GWAS data were subjected to GO analysis (Tables S3 – S5, in supplementary material, available at http://links.lww.com/PAIN/A639).

Comparing GO results obtained in GWAS and EWAS, we observed a few common functional genomic regions, defined as “neuron recognition” in BP category with shared genes: CNTNAP2, ROBO2; in CC classification in the category “neuron projection” the shared genes were CNTNAP2, ROBO2, CDH13, and GRM7. In addition, in CC classification the categories, “excitatory synapse” and “actin cytoskeleton” were also identified in both GWAS and EWAS analyses. Thus, the genomic/epigenome analyses suggested that the genomic regions associated with functions of nervous system dominate the list of the
potential candidate genes.

**5. Comparison of FI with CWP**

Since CWP and FI are highly associated, with common shared genetic factors, we were interested whether and to what extent they shared multi-omic characteristics. We have reported the results of the omics analyses of the CWP elsewhere (12,14). They were compared with the present results. Of four metabolites significantly associated with FI-score (Table 3), EAS and uridine were also significantly ($p=1.05^{-9}$ and $p=5.8^{-03}$, after adjustment for covariates) associated with CWP. Comparing the nominal significant results identified in a similar analysis design, we observed two potential common pathways: D-glutamine and D-glutamate metabolism and galactose metabolism pathways. However, they were not significant after FDR correction.

Comparing results of GWAS and EWAS implementing GO analysis for both phenotypes, we identified the two common groups of genes: (I) "Neuron recognition", with p-values ranging from 2.1$^{-2}$ (EWAS of CWP) to $p=3.0^{-3}$ (GWAS of FI-scores), and (II) "Neuron projection (terminus)" , with p-values, ranging from 2.4$^{-2}$ (GWAS of FI-scores) to $p= 1.8^{-2}$ (EWAS of CWP).

**6. Path analysis of FI and CWP**

We examined a model including the direct and indirect effect of covariates on FI scores via CWP. In other words, we hypothesized that CWP manifestation could be an independent risk factor for worsening FI-status of an individual and several studies suggest this sequence of relations between CWP and FI (29,30). First, using modified variance decomposition analysis testing the liability-threshold model of dichotomous variables (16) we examined the contribution of potential covariates (age, smoking, relative fat mass, EAS levels and leading SNPs), on CWP. Next, implementing variance decomposition analysis we estimated all possible direct and indirect effects of CWP manifestation and other covariates on FI scores.
variation. At this stage, the epigenome signals were not included in the analysis. Fig 2 summarizes the main results of path analysis showing that all tested covariates affect the CWP liability scores significantly. While age, fat mass and smoking increase the risk for CWP, EAS circulating levels decrease with raising of the CWP- scores.

Evaluating all possible direct and indirect effects on FI scores, we observed that again almost all tested covariates (CWP, age, smoking, relative fat mass but not EAS levels) exerted a significant effect on FI-scores, with clear dominance of the CWP manifestation. Remarkably, when we added C-glycosyl tryptophan, N-acetyl glycine and uridine to the analysis, (identified as independently associated with FI-scores (Table 3), they contributed their independent association to FI (Fig 2) while not altering other parameter estimates, and their own regression coefficients were virtually the same as reported in Table 3.

Discussion

As the modern human population is ageing the prevalence of frailty is increasing. Yet, the specific manifestation of frailty in any individual at a particular age varies tremendously, as does prevalence of frailty among different communities (e.g. 25). It is therefore imperative to clarify the main risk factors for incident frailty as well as its deterioration. Previous studies, including ours, have shown a significant contribution of genetic factors to FI (6,15,33), along with other strong risk factors, specifically CWP (e.g. 28,29). CWP in turn has a significant genetic component, which exerts a pleiotropic genetic effect on FI (15). The main aim of this study was to clarify the molecular-genetic nature of FI heritability, and its correlation with CWP.

OMICS analyses identified 20 top metabolites associated with FI after correction for multiple testing (p<0.0002, Table 2). However, the metabolites themselves are highly correlated and final multiple regression analysis revealed only four independently associated metabolites:
EAS, C-glycosyl tryptophan, N-acetyl glycine and uridine (Table 3). Although they represent different facets of human physiology, they appear relevant in view of the results obtained in present GWAS and EWAS of this sample, which also suggest involvement of genomic regions associated with the nervous system. Path analysis showed that the latter three metabolites were independently associated with frailty, while the effect of EAS appeared to be mediated via CWP. EAS circulating levels showed no direct path correlation with FI (Fig 2), but was highly significantly associated with CWP, which in turn was strongly related to FI.

EAS is a major precursor of testosterone and estradiol and a potential neurosteroid (https://pubchem.ncbi.nlm.nih.gov/compound/epiandrosterone). In addition, EAS is involved in blood pressure regulation (via inhibition the pentose phosphate pathway) and several other components of blood biochemistry, thus affecting blood circulation in the microvasculature.

In our dataset, unpublished analysis has also identified this metabolite to be associated with depression and anxiety. A causal role of CWP for FI has been suggested repeatedly in the literature in samples of diverse ethnicity (e.g. 29,30), however, no clear potential mechanism of association was proposed. Our previous studies suggested involvement of neurological pathways in aetiology of CWP (14), and showed that its appearance significantly correlates with neuropathic pain features (20), and with fatigue and depression (3,9). This study further suggests that steroid pathways are involved in the mechanism of interaction between frailty and pain.

The other metabolites which were related to frailty independently of CWP also point to the importance of neuro-endocrine mechanisms in frailty. Thus, tryptophan metabolism is critical to the biosynthetic pathway generating serotonin (5-hydroxytryptamin) (17; http://themedicalbiochemistrypage.org/nerves.html#5ht), a major neurotransmitter in autonomic nervous system as well as in the CNS. Its function relates to mood, cognition
(memory and learning), the regulation of appetite, sleep and others. In an earlier study from our group, C-gly Trp was also associated with age (18), which is correlated with frailty. It is likely that uridine, a component of RNA, may have a synergistic effect with serotonin on brain function by modulating serotonin release (10). Some reports have indicated that uridine modulates sleeping patterns, its administration may affect the course of mental disorder, improve memory function and pain (7). We have previously found uridine to associate with arterial stiffness in TwinsUK, (19) and with milk intake (21). Also, circulating uridine correlates significantly with the gene-expression levels of the purinergic receptor P2RY2 (19). N-acetyl glycine also fits the hypothesis of FI worsening association with possible deterioration of nervous system functioning. This enzyme is involved in the degradation of N-acylated proteins, and individuals with N-acetyl glycine deficiency will experience multiple neurological phenomena, e.g. convulsions, hearing loss and difficulty feeding (Human Metabolome www.hmdb.ca/metabolites/HMDB0000532). Thus all three molecules, appear relevant as potential molecular risk factors for FI development and progression. This conclusion is in agreement with our genomic and epigenome analysis. Although association results observed in both analyses did not reach genome-wide significance, the enrichment analysis of the nominally significant results clearly suggest prevalent association with genomic regions involved in NS functions, such as for example: “neuron recognition”, “neuron projection” and “excitatory synapse”.

Overall, our data consistently point to the association of neurological pathways markers with FI-scores progression. The association between chronic pain and frailty may be mediated by alterations in sex hormone metabolism.
Conflict of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Arthritis Research UK grant number 20682 to FW, by the MRC AimHY (MR/M016560/1) project grant to CM, and by the Israel Science Foundation grant number 1018/13 to GL. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London.

Ethics approval

This study received ethics approval from the St. Thomas’ Hospital Research Ethics Committee.

References (correct list)


22. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure


Legends to Figures

Figure 1. General outline of the study design.

Figure 2. Path analysis of liability to CWP and its potential affect on FI-score variation in relation to several common covariates. The main hypothesis is that CWP-manifestation causes FI-deterioration.
Table 1. Basic descriptive statistics of the study sample from TwinsUK

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skew±SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI-scores</td>
<td>3626</td>
<td>0.23</td>
<td>0.14</td>
<td>0.01</td>
<td>0.81</td>
<td>1.010±0.041</td>
</tr>
<tr>
<td>Age, years</td>
<td>3613</td>
<td>60.50</td>
<td>13.87</td>
<td>16.37</td>
<td>93.01</td>
<td>-0.700±0.041</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>2863</td>
<td>26.04</td>
<td>4.90</td>
<td>15.88</td>
<td>52.71</td>
<td>1.096±0.046</td>
</tr>
<tr>
<td>FAT, kg</td>
<td>2276</td>
<td>25.08</td>
<td>8.53</td>
<td>5.67</td>
<td>63.77</td>
<td>0.814±0.051</td>
</tr>
</tbody>
</table>
Table 2. Association of frailty with metabolites

<table>
<thead>
<tr>
<th>#</th>
<th>Metabolite</th>
<th>R</th>
<th>P</th>
<th>N</th>
<th>Brief description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glutamate</td>
<td>0.161</td>
<td>4.0E-16</td>
<td>2529</td>
<td>Amino acid. Glutamate metabolism</td>
</tr>
<tr>
<td>2</td>
<td>Urate</td>
<td>0.137</td>
<td>4.1E-12</td>
<td>2530</td>
<td>Nucleotide. Purine metabolism, urate metabolism</td>
</tr>
<tr>
<td>3</td>
<td>N-acetylglucose*</td>
<td>-0.137</td>
<td>2.4E-11</td>
<td>2358</td>
<td>Amino acid. Glycine, serine and threonine metabolism</td>
</tr>
<tr>
<td>4</td>
<td>C-glycosyltryptophan*</td>
<td>0.130</td>
<td>4.9E-11</td>
<td>2525</td>
<td>Amino acid. Tryptophan metabolism</td>
</tr>
<tr>
<td>5</td>
<td>Pseudouridine</td>
<td>0.123</td>
<td>6.3E-10</td>
<td>2526</td>
<td>Nucleotide. Pyrimidine metabolism, uracil containing</td>
</tr>
<tr>
<td>6</td>
<td>Docosahexaenoate (DHA; 22:6n3)</td>
<td>-0.119</td>
<td>2.0E-09</td>
<td>2530</td>
<td>Lipid. Essential fatty acid</td>
</tr>
<tr>
<td>7</td>
<td>Mannose</td>
<td>0.116</td>
<td>5.2E-09</td>
<td>2530</td>
<td>Carbohydrate. Fructose, mannose, galactose, starch, and sucrose metabolism</td>
</tr>
<tr>
<td>8</td>
<td>HWESASXX</td>
<td>0.111</td>
<td>2.7E-08</td>
<td>2512</td>
<td>Peptide. Polypeptide</td>
</tr>
<tr>
<td>9</td>
<td>Uridine*</td>
<td>-0.109</td>
<td>3.6E-08</td>
<td>2530</td>
<td>Nucleotide. Pyrimidine metabolism, uracil containing</td>
</tr>
<tr>
<td>10</td>
<td>Epiandrosterone sulphate*</td>
<td>-0.107</td>
<td>7.8E-08</td>
<td>2512</td>
<td>Lipid. Sterol/Steroid</td>
</tr>
<tr>
<td>11</td>
<td>Proline</td>
<td>0.107</td>
<td>7.8E-08</td>
<td>2530</td>
<td>Amino acid. Urea cycle; arginine-, proline- metabolism</td>
</tr>
<tr>
<td>12</td>
<td>Indolepropionate*</td>
<td>-0.106</td>
<td>8.5E-08</td>
<td>2529</td>
<td>Amino acid. Tryptophan metabolism</td>
</tr>
<tr>
<td>13</td>
<td>1-docosahexaeno-glycerophosphocholine*</td>
<td>-0.105</td>
<td>1.2E-07</td>
<td>2521</td>
<td>Lipid. Lysolipid</td>
</tr>
<tr>
<td>14</td>
<td>Gamma-glutamylvaline</td>
<td>0.105</td>
<td>1.3E-07</td>
<td>2518</td>
<td>Peptide, gamma-glutamyl</td>
</tr>
<tr>
<td>15</td>
<td>Gamma-glutamylleucine</td>
<td>0.102</td>
<td>2.7E-07</td>
<td>2529</td>
<td>Peptide, gamma-glutamyl</td>
</tr>
<tr>
<td>16</td>
<td>Gamma-glutamylphenylalanine</td>
<td>0.102</td>
<td>3.2E-07</td>
<td>2506</td>
<td>Peptide, gamma-glutamyl</td>
</tr>
<tr>
<td>17</td>
<td>N-acetylaspartine</td>
<td>0.099</td>
<td>7.0E-07</td>
<td>2526</td>
<td>Amino acid. Alanine and aspartate metabolism</td>
</tr>
<tr>
<td>18</td>
<td>Butyrylcarnitine</td>
<td>0.095</td>
<td>1.6E-06</td>
<td>2530</td>
<td>Lipid. Fatty acid metabolism (also BCAA metabolism)</td>
</tr>
<tr>
<td>19</td>
<td>Glycerol</td>
<td>0.094</td>
<td>2.0E-06</td>
<td>2530</td>
<td>Lipid. Glycerolipid metabolism</td>
</tr>
<tr>
<td>20</td>
<td>2-dimoleoylglycerophosphocholine</td>
<td>-0.104</td>
<td>2.1E-06</td>
<td>2074</td>
<td>Lipid. Lysolipid</td>
</tr>
</tbody>
</table>

Legend to table. Correlation of FI adjusted for age with the 20 selected metabolite circulating levels in total available sample. The compounds marked by stars(*) were independently significantly associated with FI-scores.
Table 3. Risk factors for frailty. Multivariable regression of FI-scores on the most highly associated metabolites, with adjustment for age, relative fat mass and smoking.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Beta(^1) ± SE(_{\text{Beta}})</th>
<th>T–test</th>
<th>P</th>
<th>Beta(^2) ± SE(_{\text{Beta}})</th>
<th>LRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epiandrosterone sulfate</td>
<td>-0.092 ± 0.019</td>
<td>-4.744</td>
<td>2.23E-06</td>
<td>-0.112 ± 0.019</td>
<td>7.7E-09</td>
</tr>
<tr>
<td>C-glycosyl tryptophan</td>
<td>0.092 ± 0.021</td>
<td>4.420</td>
<td>1.04E-05</td>
<td>0.077 ± 0.021</td>
<td>2.0E-04</td>
</tr>
<tr>
<td>N-acetyl glycine</td>
<td>-0.072 ± 0.019</td>
<td>-3.774</td>
<td>1.65E-04</td>
<td>-0.077 ± 0.020</td>
<td>1.5E-04</td>
</tr>
<tr>
<td>Uridine</td>
<td>-0.070 ± 0.019</td>
<td>-3.712</td>
<td>2.11E-04</td>
<td>-0.062 ± 0.020</td>
<td>2.45E-03</td>
</tr>
<tr>
<td>Indolepropionate</td>
<td>-0.042 ± 0.019</td>
<td>-2.223</td>
<td>3.6E-03</td>
<td>-0.015 ± 0.021</td>
<td>4.8E-01 (ns)</td>
</tr>
<tr>
<td>Age</td>
<td>0.299 ± 0.021</td>
<td>14.338</td>
<td>1.40E-44</td>
<td>0.340 ± 0.027</td>
<td>1.6E-34</td>
</tr>
<tr>
<td>Fat/HT(^2)</td>
<td>0.218 ± 0.020</td>
<td>10.655</td>
<td>7.17E-26</td>
<td>0.236 ± 0.020</td>
<td>4.2E-31</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.081±0.019</td>
<td>4.379</td>
<td>1.25E-05</td>
<td>0.083 ± 0.018</td>
<td>6.6E-06</td>
</tr>
<tr>
<td>Heritability, h(^2)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.379 ± 0.125</td>
<td>2.4E-03</td>
</tr>
</tbody>
</table>

Legends to table. Adjusted R\(^2\)= 0.260, F(8.2153)=96.08, p<0.00001; N=2162. Beta, T and p describe the effect of each variable on FI-score variation adjusted for all the others. Beta\(^1\) – regression estimates obtained in preliminary multiple regression analysis. Beta\(^2\) – regression estimates obtained in final multiple regression analysis, taken into account familial composition of the sample, and estimating heritability of the FI-scores, adjusted for all tested covariates. LRT – p-value, likelihood ratio test assuming no effect of the selected covariate, ns =p>0.05
FI participants’ recruitment (3,626)

Questionnaire & body composition assessment

OMICS analysis

- Fifty pairs of discordant twins
- Main sample 2,482 individuals
- Genome study, 2.5 mln SNP
- Metabolome study, 510 metabolites
- Epigenome study, 11 mln methylation bins

Statistical & Bioinformatics Analyses

Comparative Bioinformatics Analysis of FI and CWP
Liability to CWP

Age

Smoking

EAS

FAT/H²

SNP

Frailty score

Liability to CWP

0.111±0.042

0.230±0.028

0.275±0.015

0.150±0.050

0.204±0.033

0.192±0.031

0.116±0.037

0.771±0.063

0.199±0.019

C-glycosyl tryptophan

N-acetyl glycine

Uridine